

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2003-085169

(43)Date of publication of application : 20.03.2003

(51)Int.Cl.

G06F 17/28

(21)Application number : 2001-271907

(71)Applicant : PATOLIS CORP

(22)Date of filing : 07.09.2001

(72)Inventor : ISHIKAWA TETSUYA

FUJII ATSUSHI

HIGUCHI SHIGETO

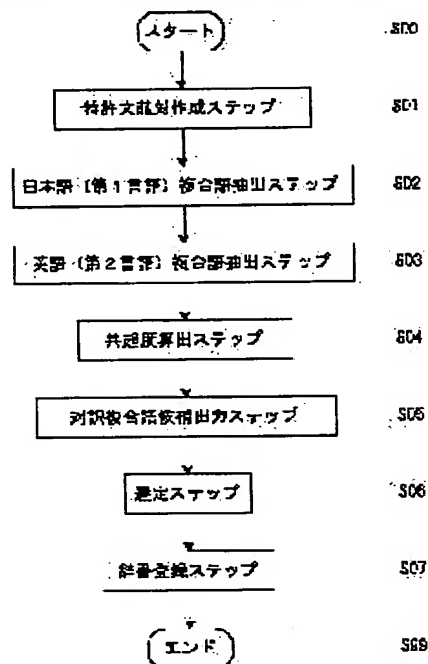
FUKUI MASATOSHI

(54) PARALLEL TRANSLATION COMPOUND WORD DICTIONARY UPDATING METHOD

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a parallel translation compound word dictionary updating method for registering a dictionary by extracting a parallel translation compound word at an inexpensive operation cost.

SOLUTION: A patent document pair file is made by gathering a pair of patent documents coincident in a priority number of a Japanese patent document and a filing number of an English patent document (S01). A Japanese compound word appearing in Japanese patent documents is extracted from the respective Japanese patent documents in the patent document pair file (S02). An English compound word appearing in English patent documents is extracted from the respective English patent documents in the patent document pair file (S03). A cooccurrence degree of respective English compound words is calculated to this Japanese compound word with every Japanese compound word in the same patent document pairs (S04). The English compound word having a high cooccurrence degree to the Japanese compound word is outputted b the proper number as parallel translation compound word candidates with every Japanese compound word (S05). A parallel translation English compound word of the Japanese compound word is selected from the parallel translation compound word candidates of the English compound word outputted with every Japanese compound word (S06). A parallel translation compound word dictionary is updated thereby (S07).



LEGAL STATUS

[Date of request for examination] 29.07.2004

[Date of sending the examiner's decision of rejection] 12.04.2005

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

BEST AVAILABLE COPY

JT5827-1-PC7
LSP 3/13/1

(19) 日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11) 特許出願公開番号
特開2003-85169
(P2003-85169A)

(43) 公開日 平成15年3月20日 (2003.3.20)

(51) Int.Cl.
G 0 6 F 17/28

識別記号

FI
G 0 6 F 17/28

テーマコード(参考)
C 5 B 0 9 1

審査請求 未請求 請求項の数10 OL (全 8 頁)

(21) 出願番号 特願2001-271907(P2001-271907)

(22) 出願日 平成13年9月7日 (2001.9.7)

(71) 出願人 501305453

株式会社 バトリス
東京都江東区塩浜2丁目4番29号

(72) 発明者 石川 徹也

茨城県つくば市春日1-2 図書館情報大
学内

(72) 発明者 藤井 敦

茨城県つくば市春日1-2 図書館情報大
学内

(74) 代理人 100108730

弁理士 天野 正景 (外1名)

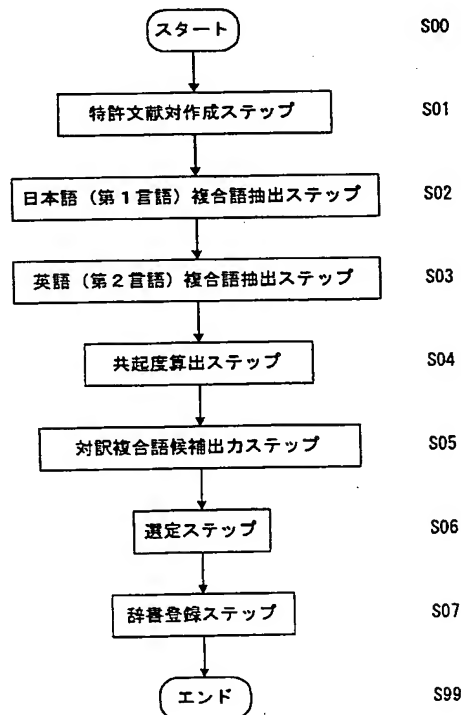
最終頁に続く

(54) 【発明の名称】 対訳複合語辞書更新方法

(57) 【要約】

【課題】 安価な運用費用でもって対訳複合語を抽出して辞書登録するための対訳複合語辞書更新方法を提供することを課題とするものである。

【解決手段】 日本語特許文献の優先権番号と英語特許文献の出願番号とが一致する一対の特許文献対を収集した特許文献対ファイルを作成し (S01)、特許文献対ファイル内のそれぞれの日本語特許文献から、この中に出現する日本語複合語を抽出し (S02)、特許文献対ファイル内のそれぞれの英語特許文献から、この中に出現する英語複合語を抽出し (S03)、同じ特許文献対において、日本語複合語毎に、この日本語複合語に対し英語複合語のそれぞれが共起する度合いを算出する (S04)。日本語複合語毎に、これに対する共起度が高い英語複合語を対訳複合語候補として適宜の数だけ出力し (S05)、日本語複合語毎に出力された英語複合語の対訳複合語候補の中から、日本語複合語の対訳英語複合語を選定し (S06)、これにより対訳複合語辞書を更新 (S07) する。



【特許請求の範囲】

【請求項 1】 第 1 言語特許文献の優先権番号と第 2 言語特許文献の出願番号とが一致する一対の特許文献対を収集した特許文献対ファイルを作成するための特許文献対作成ステップと、

上記特許文献対ファイル内のそれぞれの第 1 言語特許文献から、この中に出現する第 1 言語複合語を抽出するための第 1 言語複合語抽出ステップと、

上記特許文献対ファイル内のそれぞれの第 2 言語特許文献から、この中に出現する第 2 言語複合語を抽出するための第 2 言語複合語抽出ステップと、

同じ特許文献対において、上記第 1 言語複合語毎に、この第 1 言語複合語に対し上記第 2 言語複合語のそれぞれが共起する度合いを算出するための共起度算出ステップと、

上記第 1 言語複合語毎に、これに対する上記共起度が高い第 2 言語複合語を対訳複合語候補として適宜の数だけ出力するための対訳複合語候補出力ステップと、

上記対訳複合候補出力ステップにおいて第 1 言語複合語毎に出力された第 2 言語複合語の対訳複合語候補の中から、第 1 言語複合語の対訳第 2 言語複合語を選定するための選定ステップと、

上記選定ステップにおいて選定された第 1 言語複合語とその対訳第 2 言語複合語を対訳複合語辞書に登録するための辞書登録ステップとを備えていることを特徴とする対訳複合語辞書更新方法。

【請求項 2】 請求項 1 に記載された対訳複合語辞書更新方法において、

上記特許文献対作成ステップにおいて作成される特許文献対ファイルは、上記第 1 言語特許文献と上記第 2 言語特許文献とにおける互いに対応する一部の欄を用いて作成された特許文献対ファイルであることを特徴とする対訳複合語辞書更新方法。

【請求項 3】 請求項 2 に記載された対訳複合語辞書更新方法において、

上記一部の欄は、発明の名称の欄、要約の欄、特許請求の範囲の欄、発明の詳細な説明の欄あるいは実施例の欄の中の一つ欄又は複数の欄であることを特徴とする対訳複合語辞書更新方法。

【請求項 4】 請求項 1 から請求項 3 までのいずれかに記載された対訳複合語辞書更新方法において、

上記対訳複合語候補出力ステップでは、現在の対訳複合語辞書に未登録の対訳複合語が抽出されて出力されることを特徴とする対訳複合語辞書更新方法。

【請求項 5】 請求項 1 から請求項 4 までのいずれかに記載された対訳複合語辞書更新方法において、

上記特許文献対作成ステップにおいて作成される特許文献対ファイルは、各言語毎のファイルの組からなることを特徴とする対訳複合語辞書更新方法。

【請求項 6】 請求項 1 から請求項 5 までのいずれかに

記載された対訳複合語辞書更新方法において、

上記第 1 言語複合語抽出ステップにおける抽出される第 1 言語複合語及び上記第 2 言語複合語抽出ステップにおける抽出される第 2 言語複合語を、各特許文献を形態素解析して抽出された単語を予め定められた結合規則に則って生成することを特徴とする対訳複合語辞書更新方法。

【請求項 7】 請求項 1 から請求項 6 までのいずれかに記載された対訳複合語辞書更新方法において、

上記共起度 $S(J, E)$ には、

$$\text{式 } S(J, E) = \log F(J, E) \cdot 2 F(J, E) / (F(J) + F(E))$$

(ここで、 J 、 E はそれぞれ第 1 言語複合語及び第 2 言語複合語を表す。また、 $F(J)$ 及び $F(E)$ は、それぞれ第 1 言語特許文献における第 1 言語複合語の出現頻度、及び、第 2 言語特許文献における第 2 言語複合語の出現頻度を表しており、 $F(J, E)$ は、特許文献対に第 1 言語複合語と第 2 言語複合語とが共に現れる共出現頻度である。) が使用されることを特徴とする対訳複合語辞書更新方法。

【請求項 8】 請求項 1 から請求項 7 までのいずれかに記載された対訳複合語辞書更新方法において、

上記第 1 言語及び第 2 言語は、それぞれ日本語及び英語であり、

上記第 1 言語特許文献、及び、上記第 2 言語特許文献は、それぞれ日本特許文献、及び、米国特許文献であることを特徴とする対訳複合語辞書更新方法。

【請求項 9】 請求項 8 に記載された対訳複合語辞書更新方法において、

上記結合規則における日本語複合語を生成するための結合規則は、

1 非自立名詞、代名詞、数詞を除いて、文中の名詞、未知語、接頭詞、自立動詞、及び自立形容詞の連続を結合する。

2 結合した末尾の単語が接頭詞あるいは自立形容詞の場合は複合語としない。

3 結合した単語の数が 3 より多い複合語については、このほかにこれより少ない単語の結合を生成する。

4 接尾名詞は前の単語と結合する。

5 片仮名未知語については一つの単語として扱って上記規則を適用する。

こととする。また、英語複合語を生成するための結合規則は、

1 形容詞及び名詞の連続を結合する。

2 結合した単語の数が 3 より多い複合語については、このほかにこれより少ない単語の結合を生成する。

3 結合した末尾の単語が自立形容詞の場合は複合語としない。こととする。

ことを特徴とする対訳複合語辞書更新方法。

【請求項 10】 請求項 1 から請求項 9 までのいずれかに記載された対訳複合語辞書更新方法において、

上記第 1 言語特許文献は、検索によって予め技術分野毎に分けられたものであることを特徴とする対訳複合語辞書更新方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、複数言語間の言語処理技術の分野に関し、安価な運用費用でもって対訳複合語を抽出して辞書登録するための対訳複合語辞書更新方法に関する。

【0002】

【従来の技術】インターネットの普及によって様々な外国語情報を容易に入手できる時代になり、それらを母国語で読みたい、また逆に自分の情報を他言語ユーザにも読んでほしいという要求が増加している。しかし、インターネット上の情報（Web ページなど）には、日々新語が登場し、既存の翻訳辞書に登録されていない新語や専門用語などが数多く含まれているため、機械翻訳や多言語検索で対応できないことが多い。この未登録語を如何に迅速にしかも安価に辞書登録するかが言語処理技術の上で重要な課題となっている。

【0003】このような問題への対処として、対訳関係にある多言語コーパス（例文集）から単語や句の単位で対訳を自動抽出する統計的手法が考えられているが、新語の発生に追従して最新のコーパスを迅速に入手する見通しがないため、実用化されるに至っていない。

【0004】Web から対訳関係にあるページや対訳情報を掲載したページを抽出できるなら、定期的にこれを収集することにより安価な対訳コーパスを取得することができる。しかしながら、Web ページには低品質な情報が混在しているため、上質な情報だけを機械的に取得することが困難であり、日本語ページには、ヨーロッパ言語間のように決まって対訳があるようなページが少なく、更に、複合語の場合、単に単語ごとの対訳単語を複合させただけでは十分ではない。

【0005】

【発明が解決しようとする課題】本発明は、このような実状に鑑み、安価な運用費用でもって対訳複合語を抽出して辞書登録するための対訳複合語辞書更新方法を提供することを課題とするものである。

【0006】

【課題を解決するための手段】上記課題は以下の手段により解決される。すなわち、第 1 番目の発明の解決手段は、第 1 言語特許文献の優先権番号と第 2 言語特許文献の出願番号とが一致する一対の特許文献対を収集した特許文献対ファイルを作成するための特許文献対作成ステップと、上記特許文献対ファイル内のそれぞれの第 1 言語特許文献から、この中に出現する第 1 言語複合語を抽出するための第 1 言語複合語抽出ステップと、上記特許文献対ファイル内のそれぞれの第 2 言語特許文献から、この中に出現する第 2 言語複合語を抽出するための第 2

言語複合語抽出ステップと、同じ特許文献対において、上記第 1 言語複合語毎に、この第 1 言語複合語に対し上記第 2 言語複合語のそれぞれが共起する度合いを算出するための共起度算出ステップと、上記第 1 言語複合語毎に、これに対する上記共起度が高い第 2 言語複合語を対訳複合語候補として適宜の数だけ出力するための対訳複合語候補出力ステップと、上記対訳複合候補出力ステップにおいて第 1 言語複合語毎に出力された第 2 言語複合語の対訳複合語候補の中から、第 1 言語複合語の対訳第 2 言語複合語を選定するための選定ステップと、上記選定ステップにおいて選定された第 1 言語複合語とその対訳第 2 言語複合語を対訳複合語辞書に登録するための辞書登録ステップとを備えている対訳複合語辞書更新方法である。

【0007】第 2 番目の発明の解決手段は、第 1 番目の発明の対訳複合語辞書更新方法において、上記特許文献対作成ステップにおいて作成される特許文献対ファイルを、上記第 1 言語特許文献と上記第 2 言語特許文献とにおける互に対応する一部の欄を用いて作成された特許文献対ファイルとしたものである。

【0008】第 3 番目の発明の解決手段は、第 2 番目の発明の対訳複合語辞書更新方法において、上記一部の欄を、発明の名称の欄、要約の欄、特許請求の範囲の欄、発明の詳細な説明の欄あるいは実施例の欄の中の一つ欄又は複数の欄としたものである。

【0009】第 4 番目の発明の解決手段は、第 1 番目から第 3 番目までの発明の対訳複合語辞書更新方法において、上記対訳複合語候補出力ステップでは、現在の対訳複合語辞書に未登録の対訳複合語が抽出されて出力されるものである。

【0010】第 5 番目の発明の解決手段は、第 1 番目から第 4 番目までの発明の対訳複合語辞書更新方法において、上記特許文献対作成ステップにおいて作成される特許文献対ファイルを、各言語毎のファイルの組からなるようにしたものである。

【0011】第 6 番目の発明の解決手段は、第 1 番目から第 5 番目までの発明の対訳複合語辞書更新方法において、上記第 1 言語複合語抽出ステップにおける抽出される第 1 言語複合語及び上記第 2 言語複合語抽出ステップにおける抽出される第 2 言語複合語を、各特許文献を形態素解析して抽出された単語を予め定められた結合規則に則って生成するようにしたものである。

【0012】第 7 番目の発明の解決手段は、第 1 番目から第 6 番目までの発明の対訳複合語辞書更新方法において、上記共起度 $S(J, E)$ には、

$$\text{式 } S(J, E) = \log F(J, E) \cdot 2 F(J, E) / (F(J) + F(E))$$

（ここで、 J 、 E はそれぞれ第 1 言語複合語及び第 2 言語複合語を表す。また、 $F(J)$ 及び $F(E)$ は、それぞれ第 1 言語特許文献における第 1 言語複合語の出現頻度、及び、第 2 言語特許文献における第 2 言語複合語の出現頻

度を表しており、F(J,E)は、特許文献対に第1言語複合語と第2言語複合語とが共に現れる共出現頻度である。)を使用したものである。

【0013】第8番目の発明の解決手段は、第1番目から第7番目までの発明の対訳複合語辞書更新方法において、上記第1言語及び第2言語を、それぞれ日本語及び英語とし、上記第1言語特許文献、及び、上記第2言語特許文献を、それぞれ日本特許文献、及び、米国特許文献としたものである。

【0014】第9番目の発明の解決手段は、第8番目の発明の対訳複合語辞書更新方法において、上記結合規則における日本語複合語を生成するための結合規則を、1 非自立名詞、代名詞、数詞を除いて、文中の名詞、未知語、接頭詞、自立動詞、及び自立形容詞の連続を結合する。2 結合した末尾の単語が接頭詞あるいは自立形容詞の場合は複合語としない。3 結合した単語の数が3より多い複合語については、このほかにこれより少ない単語の結合を生成する。4 接尾名詞は前の単語と結合する。5 片仮名未知語については一つの単語として扱って上記規則を適用する。こととし、また、英語複合語を生成するための結合規則は、1 形容詞及び名詞の連続を結合する。2 結合した単語の数が3より多い複合語については、このほかにこれより少ない単語の結合を生成する。3 結合した末尾の単語が自立形容詞の場合は複合語としない。こととしたものである。

【0015】第10番目の発明の解決手段は、第1番目から第9番目までのいずれかの対訳複合語辞書更新方法において、上記第1言語特許文献を、検索によって予め技術分野毎に分けられたものとした対訳複合語辞書更新方法である。

【0016】

【発明の実施の形態】本発明は、対応特許文献から複合語対訳情報を抽出し、これにより辞書更新を行うので、原理的には言語の種類を問わない。そのため、ここでは日本と米国に出願された対応特許から対訳複合語辞書を更新する方法を例にして本発明を説明する。本発明では、優先権主張に基づいて日本と米国に出願された対応特許を対訳コーパスとして利用し、そこから日英対訳複合語を抽出し、対訳複合語辞書を更新する。

【0017】図1は、本発明の対訳複合語辞書更新方法における処理の流れの概念を示すフロー説明図である。この実施例では、日本の公開特許公報データベース及び米国の特許明細書データベースを用いる。ここでは、請求項中の第1言語が日本語に、第2言語が英語(米語)に、第1言語特許文献が日本公開特許公報に、第2言語特許文献が米国特許明細書に対応する。

【0018】処理がスタート(ステップS00)すると、特許文献対作成ステップ(S01)において、日本公開特許公報の優先権番号と米国特許明細書の出願番号とが一致する一対の特許文献対を収集した特許文献対フ

ァイルが作成される。なお、図7には日本公開特許公報と米国特許明細書とにおいて優先権番号と出願番号(点線囲み)が対応関係にあることを示す。

【0019】図2は、この特許文献対作成ステップのフロー概要説明図である。ステップS11において、日本公開特許公報データベースの中から、優先権主張番号と優先権主張国を参照することにより米国優先権を主張して出願された出願の公報を抽出し、米国優先権主張公開公報ファイルを作成する。なお、分割出願等により同じ優先権番号を持つ日本公開特許公報が複数存在する場合、その全てを対象とすることができる。

【0020】ステップS12では、この優先権主張番号をキーとして、米国特許明細書データベースから優先権番号と同じ出願番号を持つ米国特許明細書を抽出して、日本出願対応米国特許ファイルを作成する。これらの米国優先権主張公開公報ファイルと日本出願対応米国特許ファイルとは、互いに対をなすので特許文献対ファイルとして扱うが、これは日米公報の対応がとれば単一のファイルであっても別のファイルであってもよい。

【0021】日本語(第1言語)複合語抽出ステップS02において、特許文献対ファイル内のそれぞれの日本公開特許公報から、この中に出現する日本語(第1言語)複合語を抽出する。図3は日本語(第1言語)複合語抽出ステップの内容を示すフロー概要説明図である。

【0022】ステップS01で作成した米国優先権主張公開公報ファイルから形態素解析(ステップS21)によって単語を抽出する。なお、この単語の抽出には、形態素解析・品詞付与プログラム、例えば、「茶釜」(ちゃせん、奈良先端科学技術大学院大学情報科学研究科自然言語処理学講座(松本研究室)、<http://chasen.aist-nara.ac.jp/chasen/distribution.html.ja>)、を用いることができる。また、特許公報に非常に多く見られる「該システム、本発明、各プロセス、前記処理」中の「該、本、各、前記」等の接頭語は接辞処理によって除去する。

【0023】抽出された単語から日本語複合語を結合するための結合規則を設定するため、対訳コーパス(特許文献対ファイル)から抽出する日本語複合語の結合規則を以下のように設定する。

【0024】1 非自立名詞、代名詞、数詞を除いて、文中の名詞、未知語、接頭詞、自立動詞、及び自立形容詞の連続を結合する。この例を、図8に示す。

2 結合した末尾の単語が接頭詞あるいは自立形容詞の場合は複合語としない。この例を図9に示す。

3 結合した単語の数が3より多い複合語については、このほかにこれより少ない単語の結合を生成する。この例を図10に示す。

4 接尾名詞は前の単語と結合する。

5 片仮名未知語については一つの単語として扱い、以上の基準を適用する。この例を図11に示す。

【0025】以上の日本語結合規則に基づき、ステップS22においては、上記ステップS21で抽出された単語を結合し、日本語複合語ファイルを得る。

【0026】次に、ステップS03において、特許文献対ファイル内のそれぞれの米国特許明細書から、この中に出現する英語（第2言語）複合語を抽出する。図4は英語（第2言語）複合語抽出ステップの内容を示すフロー概要説明図である。

【0027】ステップS01で作成した日本出願対応米国特許ファイルから品詞付与プログラム（ステップS31）によって単語を抽出する。なお、品詞付与プログラムには、例えば「Brill Tagger」（<http://www.cs.jhu.edu/~brill/home.html>）を用いることができる。

【0028】抽出された単語から英語複合語を結合するための英語結合規則を設定するために、英語複合語の英語結合規則を以下のように設定する。

【0029】1 形容詞及び名詞の連続を結合する。この例を図12に示す。

2 結合した単語の数が3より多い複合語については、このほかにこれより少ない単語の結合を生成する。

3 結合した末尾の単語が自立形容詞の場合は複合語としない。この例を図13に示す。

なお、英語の場合等位接続詞(for, in, ofなど)による倒置複合語があるが専門用語辞書中に等位接続詞を含む用語は少ないため、この例では等位接続詞を意識した結合は行わない。

【0030】この英語結合規則に基づき、ステップS32においては、上記ステップS31で抽出された単語を結合し、英語複合語ファイルを作成する。

【0031】上記複合語ファイルでは、日本語複合語と上記英語複合語とは未だ対応づけられていない。この対応づけのために、共起度算出ステップS04及び対訳複合語候補出力ステップS05が実行される。この共起度算出ステップS04と対訳複合語候補出力ステップS05の内容を図5に示す。

【0032】共起度算出ステップS04内のステップS41では、同じ特許文献対において、日本語（第1言語）複合語毎に、この日本語（第1言語）複合語に対し英語（第2言語）複合語のそれぞれが共起する度合いが算出される。共起度は、その日本語複合語と英語複合語が平行して現れる程度を示す尺度であって、例えば、北村美穂子氏、松本裕治氏によって「重み付きDice係数」として提案された以下のスコアS(J,E)を用いることができる（注：北村美穂子、松本裕治、対訳コーパスを利用した対訳表現の自動抽出、情報処理学会論文誌、Vol. 38, No.4, pp.727-736, 1997.）。

【0033】 $S(J, E) = \log F(J, E) \cdot 2 F(J, E) / (F(J) + F(E))$

ここで、J、Eはそれぞれ日本語（第1言語）複合語及び英語（第2言語）複合語を表す。また、F(J)及びF(E)

は、それぞれ日本語（第1言語）特許文献における日本語（第1言語）複合語の出現頻度、及び、米国（第2言語）特許文献における英語（第2言語）複合語の出現頻度を表しており、F(J,E)は、特許文献対に日本語複合語と英語複合語とが共に現れる共出現頻度である。

【0034】この式により、日本語複合語と英語複合語との様々な組み合わせについての共起度を求めることができる。ステップS42では、各日本語複合語について、共起度（スコア）の高い順に、N個（例えば4個あるいは5個）の対訳候補リストを出力する。なお、図示しないが、このとき、更新対象となる現在の対訳複合語辞書に掲載されている複合単語対は除外し、未登録の複合語対だけを対訳候補リストとして出力する。

【0035】公報上の対応箇所において、ある日本語複合語が出現するときには、ほぼ決まっている高い頻度で特定の英語複合語が出現するとすれば、そのことは、その日本語複合語とその英語複合語とが密接な関係を持ち、更には2つの複合語が互いに対訳の関係にある可能性が大きいことを示すものと言える。なお、共起度の高いものだけを適宜の数だけ出力することは、次の人手による選定作業の負担が軽減されるという意味もある。

【0036】選定ステップS06及び辞書登録ステップS07では、ステップS42で出力された対訳候補リストに基づいて、各日本語複合語に対する対訳複合語が選定される。図6に、選定ステップS06及び辞書登録ステップS07の内容を示す。ステップS51では、対訳候補リスト、つまり各日本語複合語毎に対応する複数の英語複合語（候補）、が表示（あるいはプリントアウト）され、ステップS52では、人がこれを見てその日本語複合語に対応すると判断できる英語複合語を選定する。ここに現れる日本語複合語全てについて選定が終わり、対訳選定リストが完成する。

【0037】ステップS53によって完成した対訳選定リストの内容は、古い対訳複合語辞書に追加され対訳複合辞書の更新が終了する。

【0038】実験では、日本公開特許公報として1995年～1999年の5年間に公開された約175万件の内、米国での優先権主張を伴う出願32,590件を使用した。優先権主張番号を元に米国登録公報を抽出した結果、特許公報32,896件が得られた。抽出した日本語数は延べ1,100,416語、異なり345,291語、英語数は延べ1,233,070語、異なり439,020語であった。抽出された対訳例を図14の表に示す。

【0039】対訳辞書の更新はその性格上、正解率が100%でない限り自動更新登録はできないため、どうしても人手による修正確認作業は避けられないが、人手作業の負担を大幅に軽減することができる。運用面を考慮するとこれは非常に大きなメリットである。

【0040】以上の説明では、特許文献対ファイルを作

成するための公報上の対応箇所として、発明の名称や要約を例とする場合についてあげたが、これだけでなく、特許請求の範囲、発明の詳細な説明、実施例の各欄を対象とすることもできる。更に、分野別の複合語辞書の場合、国際分類、F I、F ターム、対応 U S 分類等に基づく検索によって、予め技術分野毎に特許文献を分けた上で上記処理を行うことができる。これによって、専門分野ごとに辞書更新を行うことができる。

【0041】以上の実施例に示すように、書式の比較的整った特許公報（明細書）を利用できるので、日々新しく出現する複合語に対訳複合語辞書を対応させることができる。また、特許公報（明細書）は各国政府が提供するデータであるため安価に取得でき、安い運用費用でもって対訳複合語辞書を更新することができる。更に、新しい複合語候補は対訳として蓋然性の高いものだけが限定されて表示（あるいはプリントアウト）されるため、選定作業が容易になる。

【0042】

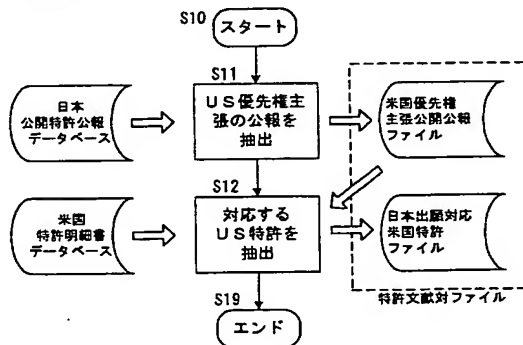
【発明の効果】本発明によれば、書式の比較的整った特許公報（明細書）を利用できるので、日々新しく出現する複合語に対訳複合語辞書を対応させることができるという効果を奏する。また、特許公報（明細書）は各国政府が提供するデータであるため安価に取得でき、安い運用費用でもって対訳複合語辞書を更新することができるという効果を奏する。更に、新しい複合語候補は対訳として蓋然性の高いものだけが限定されて表示（あるいはプリントアウト）されるため、選定作業が容易になるという効果を奏する。

【図面の簡単な説明】

【図1】本発明の対訳複合語辞書更新方法における処理の流れの概念を示すフロー説明図である。

【図2】

特許文献対作成ステップ



【図2】特許文献対作成ステップのフロー概要説明図である。

【図3】日本語（第1言語）複合語抽出ステップの内容を示すフロー概要説明図である。

【図4】英語（第2言語）複合語抽出ステップの内容を示すフロー概要説明図である。

【図5】この共起度算出ステップS04と対訳複合語候補出力ステップS05の内容を示すフロー概要説明図である。

【図6】選定ステップS06及び辞書登録ステップS07の内容を示すフロー概要説明図である。

【図7】日本公開特許公報と米国特許明細書とにおいて優先権番号と出願番号（点線囲み）が対応関係にあることを示すための説明図である。

【図8】非自立名詞、代名詞、数詞を除いて、文中の名詞、未知語、接頭詞、自立動詞、及び自立形容詞の連続を結合する例（日本語）を示す説明図である。

【図9】結合した末尾の単語が接頭詞あるいは自立形容詞の場合は複合語としない例（日本語）を示す説明図である。

【図10】結合した単語の数が3より多い複合語について、これより少ない単語の結合を生成する場合の例（日本語）を示す説明図である。

【図11】片仮名未知語について一つの単語として扱った例（日本語）を示す説明図である。

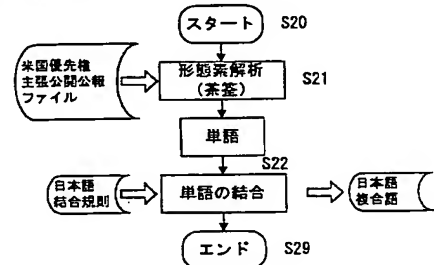
【図12】形容詞及び名詞の連続を結合する例（英語）を示す説明図である。

【図13】結合した末尾の単語が自立形容詞の場合は複合語としない例（英語）を示す説明図である。

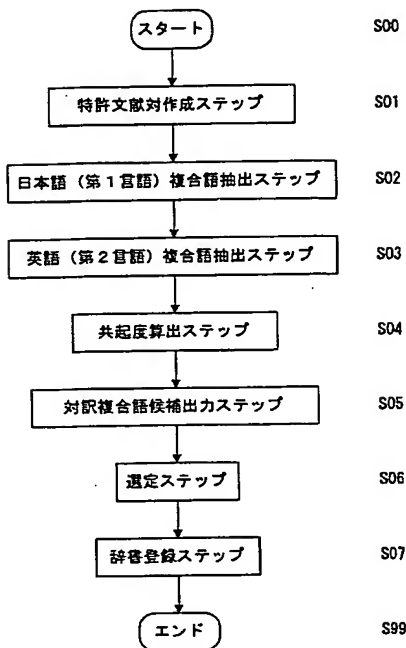
【図14】抽出された対訳例を示す表である。

【図3】

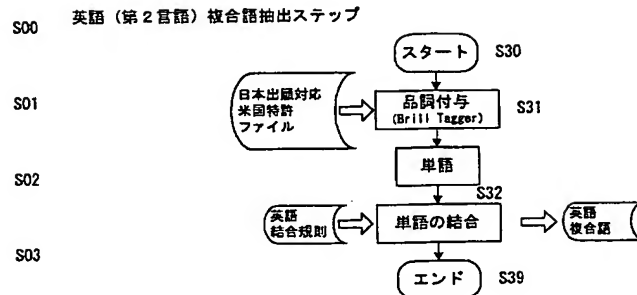
日本語（第1言語）複合語抽出ステップ



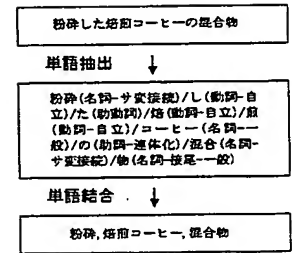
【図 1】



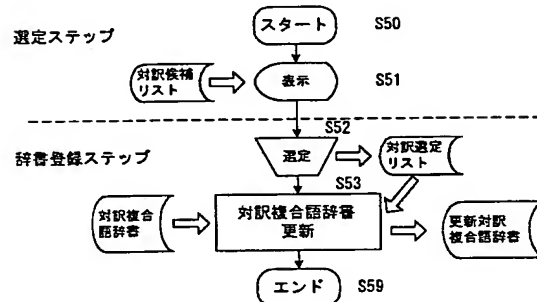
【図 4】



【図 8】

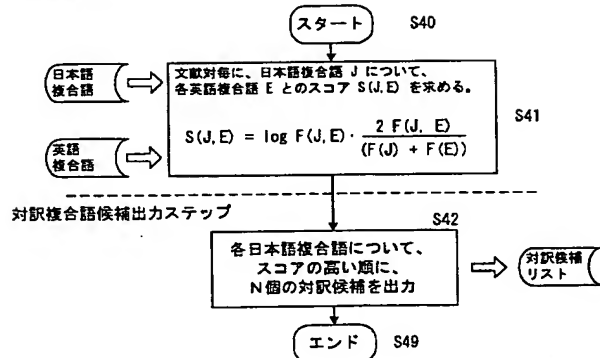


【図 6】

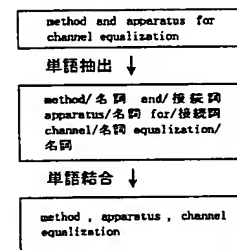


【図 5】

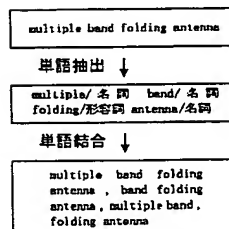
共起度算出ステップ



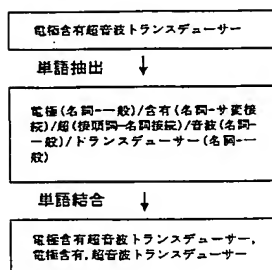
【図 12】



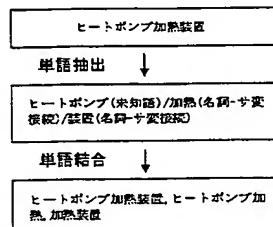
【図 13】



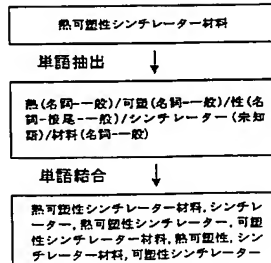
【図 9】



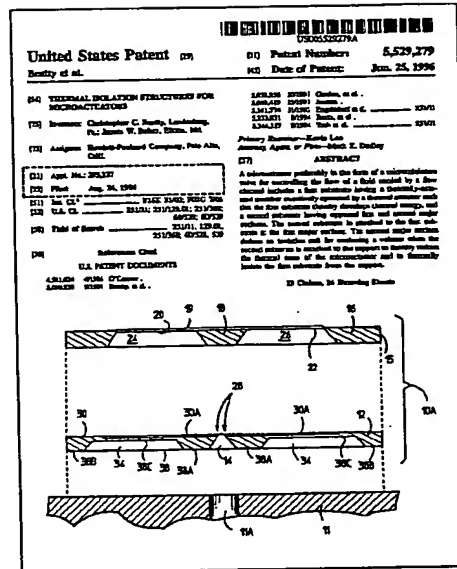
【図 10】



【図 11】



(印) 日本郵政省 (F) DD 公開特許公報 (A)	
(11) 特許庁登録番号 特開平 6-114278	
(CO) 出願日 平成 6 年 01 月 27 日	
(31) 発元 (31) [*] 7151 社 日立 G09D 7/03 H01L 21/26	発元番号 7151 社 日立 発元名称 日立 社 発元住所 東京都千代田区千代田 1-1-1 発元代表者 代表取締役社長 佐々木 隆夫
(41) 出願番号 特開平 7 年 01 月 27 日	(51) 国際分類 H01L 21/26
(61) 優先権主張 特開平 6 年 01 月 27 日	(71) 出願人 日立 社
(81) 出願国 日本	(91) 出願日 平成 6 年 01 月 27 日
(91) 出願日 平成 6 年 01 月 27 日	(92) 出願日 平成 6 年 01 月 27 日
(93) 出願日 平成 6 年 01 月 27 日	(94) 出願日 平成 6 年 01 月 27 日
(95) 出願日 平成 6 年 01 月 27 日	(96) 出願日 平成 6 年 01 月 27 日
(97) 出願日 平成 6 年 01 月 27 日	(98) 出願日 平成 6 年 01 月 27 日
(99) 出願日 平成 6 年 01 月 27 日	(100) 出願日 平成 6 年 01 月 27 日



抽出された対訳の例

アルケニル含有ポリジオルガノシロキサン	alkenyl-containing polydiorganosiloxane
イオントラップ質量スペクトロメータ	ion trap mass spectrometer
インドリルアルキルピペラジニルピリジン	indolylalkylpiperazinyl pyridine
エアバッグキャニスタ	air bag canister
ジオルガノポリシロキサンポリマー	diorganopolysiloxane polymer
シリコーン感圧接着剤組成物	silicone pressure sensitive adhesive compositions
セルローストリアセテート写真	cellulose triacetate photographic
セルローストリアセテート写真フィルムベース	cellulose triacetate photographic film base
加硫性エラストマーコンパンド	vulcanizable elastomeric compound
感熱色素転写システム	thermal dye transfer system
小型走査共焦点顕微鏡	miniature scan confocal microscope
電子マネーシステム	electronic-monetary system
蠕動ポンプ	peristaltic pump

Fターム(参考) 5B091 AA05 AB11 BA02 CC02 CC16

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.